

Architektur eines Open Source BI Systems mit Geo-Erweiterung

Anett Mehler-Bicher
Klaus B. Böhm

Veröffentlicht in:
Multikonferenz Wirtschaftsinformatik 2012
Tagungsband der MKWI 2012
Hrsg.: Dirk Christian Mattfeld; Susanne Robra-Bissantz



Braunschweig: Institut für Wirtschaftsinformatik, 2012

Architektur eines Open Source BI Systems mit Geo-Erweiterung

Anett Mehler-Bicher

Fachhochschule Mainz, Fachbereich Wirtschaft, Fachgruppe Wirtschaftsinformatik,
55128 Mainz, E-Mail: anett.bicher@fh-mainz.de

Klaus B. Böhm

Fachhochschule Mainz, Fachbereich Technik, Lehrinheit Geoinformatik & Vermessung,
55128 Mainz, E-Mail: klaus.boehm@fh-mainz.de

Abstract

In diesem Paper wird die Kombination von einer Open Source BI Lösung mit geo-bezogener Analyse und Visualisierung adressiert. Im Fokus stehen dabei die Architektur des Gesamtsystems sowie die Kommunikation der BI und Geo-Visualisierungsbezogenen Komponenten.

Ziel ist es, eine Widget-basierte Architektur als Best-Practice Ansatz vorzustellen und aufzuzeigen, welche Vorteile diese in der Nutzung bietet. Die Evaluierung des vorgestellten Ansatzes erfolgt in einem Anwendungsfall aus dem Gesundheitswesen.

1 Einleitung

Business Intelligence gewinnt in Unternehmen immer mehr an Bedeutung; auch Open Source Lösungen wie pentaho[22], palo[21] oder JasperSoft[16] haben inzwischen einen Reifegrad erlangt, der einen Einsatz in Unternehmen möglich und interessant macht.[15]

Wenig erschlossen hinsichtlich Business Intelligence– unabhängig ob kommerzielle oder Open Source Lösung –ist bislang die Verarbeitung geo-referenzierter Daten. Mehr als 80 Prozent aller Geschäftsdaten haben einen Raumbezug und bieten somit geo-referenzierte Informationen, die für die Analyse und Simulation wertvoll eingesetzt werden können.[4], [8] Bisher waren geo-basierte Analysen insbesondere eine Domäne von Geo-Informationssystemen (GIS). Die Kombination aus BI und geo-basierten Werkzeugen schafft neuartige Möglichkeiten für Analysen/Visualisierungen und Simulationen. In diesen Kontext sind die Ergebnisse, die in der Disziplin GeoVisualAnalytics[20], [25] erzielt werden, einzuordnen.

Durch den vereinfachten Zugriff auf Geo-Daten wie z. B. durch OpenStreetMap, Google Maps oder Bingmaps für raumbezogene Abfragen bei Datenbankmanagementsystemen verwischen die klassischen Branchengrenzen von GIS-Anwendungen.

Die hier vorgestellten Forschungsergebnisse, die als best-Practice Ansatz zu verstehen sind, wurden in einem vom Bundesministerium für Wirtschaft und Technologie geförderten Projekt erzielt, das den Gesundheitsmarkt fokussiert. [1]

Der Gesundheitsmarkt ist ein typischer Anwendungsbereich, der von BI Lösungen profitieren kann, derzeit aber noch selten unterstützt wird bzw. in dem der Reifegrad entsprechender BI Lösungen noch relativ gering ist. [4] Gerade aber geo-referenzierte Daten sind notwendig, um richtige Schlussfolgerungen zu ziehen. Disease Management Programme oder Einflüsse von Umweltbedingungen (z. B. Umweltverschmutzung) auf Krankheiten sind Beispiele im Gesundheitsbereich, bei denen erst die Berücksichtigung geo-referenzierter Daten aussagekräftige Resultate liefern kann. Derartige Daten können üblicherweise nicht von BI Lösungen verarbeitet werden.

Aus konzeptioneller Sicht wird daher ein System angestrebt, wie es in Bild 1 dargestellt ist.

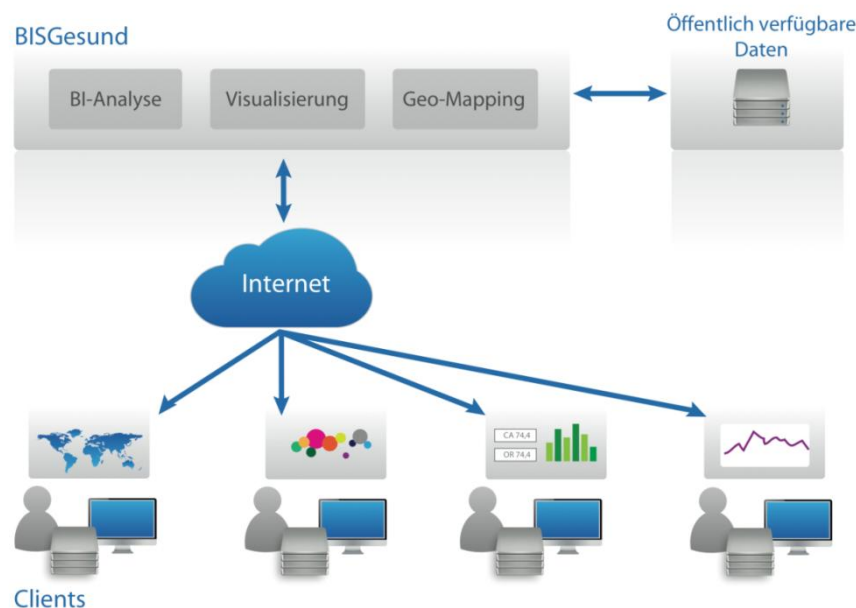


Bild 1: Konzeption eines BI Systems mit Geo-Visualisierung

Eine BI Lösung, die auch geo-referenzierte Daten verarbeiten kann, benötigt andere Darstellungsmechanismen als konventionelle Lösungen, die sich auf die Darstellung in Tabellen und Charts konzentrieren. Geo-referenzierte Daten und darauf basierende Analysen sowie Auswertungen benötigen konsequenterweise eine kartenbasierte Visualisierung.

Bei der Darstellung auf der Karte lassen sich verschiedene kartographische Techniken anwenden (vgl. Bild 2). Neben der bekannten flächenhaften Darstellung (Choroplethenkarte) wurden weitere Ansätze ausgewählt, die für die gegebene Anwendung aussagekräftige Ergebnisse liefern:

- Die Marker-Darstellung eignet sich für Informationen, die sich eindeutig auf einen Standort beziehen.[19]
- Die Cluster-Darstellung ermöglicht eine Generalisierung/Verdichtung unter Verwendung eines Fangradius.[17]
- Die Heatmap bietet sich zur Darstellung zweidimensionaler Verteilungen an.[7]

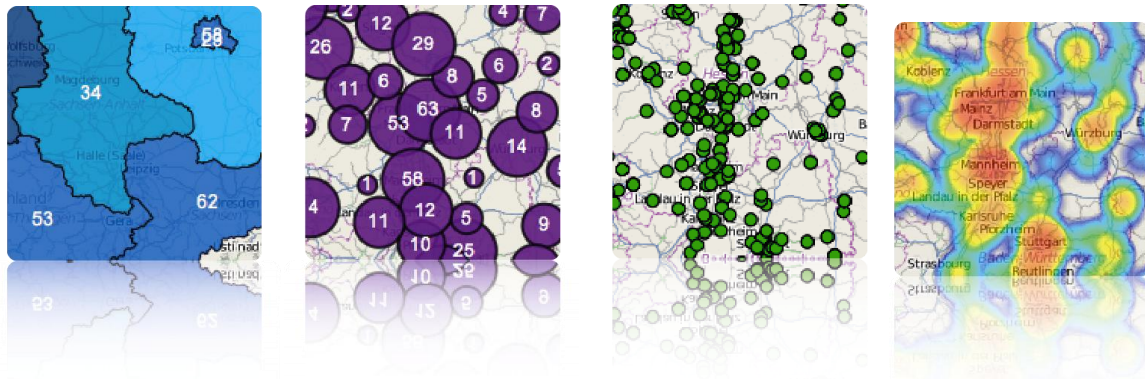


Bild 2: Kartographische Visualisierungsmöglichkeiten
 (von links: Choroplethenkarte, Cluster-Darstellung, Marker-Darstellung, Heatmap)

Das bedeutet, dass sich bei einer mit Geo-Visualisierung kombinierten BI Lösung Darstellungsformen wie Tabelle und Karte synchronisiert nutzen lassen müssen. Die Daten, die in einer Tabelle dargestellt werden, müssen bei Auswahl einer Zeile oder Spalte auf die Karte übertragen werden. Ein Zoomen in der Karte wiederum muss eine entsprechende Anpassung in der Tabelle bewirken.

Durch die Integration geo-referenzierter Daten in das BI System und die Verschmelzung von Tabellen- und Kartendarstellung ergeben sich neue Anforderungen an die Benutzerführung, aber auch an die zugrundeliegende Architektur des Systems.

Die sich daraus ergebenden Anforderungen münden in das Architekturkonzept, das im Rahmen des Drittmittelprojekts BISGesund[1] als Best Practice Lösung entwickelt wurde und in weiteren Schritten hinsichtlich eines prinzipiellen Lösungsansatzes noch validiert werden muss. Eine weitere Vorgabe des Drittmittelprojekts war, dass das System auf Basis einer Open Source BI Lösung zu realisieren ist.

2 Anforderungen

2.1 Anforderungen an BI Systeme mit Geo-Visualisierung

Mit Business Intelligence– ob klassisch oder mit Erweiterung um geo-referenzierte Daten – bezeichnet man grundsätzlich die entscheidungsorientierte Datensammlung und -aufbereitung mit der Zielsetzung, geschäftsrelevante Informationen zu analysieren und darzustellen. [4]

Erweitert man gängige BI Lösungen um geo-bezogene Daten sowie entsprechende Analyse- und Visualisierungsmöglichkeiten, erhöht die Hinzunahme von raumbezogenen Daten substantiell die Komplexität einer „Geo-BI Lösung“ durch weitere Analyse- und Visualisierungsmöglichkeiten. Eine Synchronisierung der Darstellung z. B. muss nicht nur zwischen Tabellen und Charts erfolgen, sondern zusätzlich den Aspekt Karte integrieren. Insbesondere die Fragestellung, wann, wo und wie sich geo-basierte Analyse- und Visualisierungsmöglichkeiten anbieten, führt zu weiteren Anforderungen:

- **Durchgängige Integration geo-referenzierter Daten**
Geo-bezogene Daten müssen sowohl im Data-Warehouse gespeichert als auch den zur Verfügung stehenden Tools zwecks Reporting, Analyse, Forecasting oder Data Mining zur Auswertung zur Verfügung gestellt werden. Die raum- oder geo-bezogenen Daten sind durchgängig einzubinden. Weiterhin ist die Verarbeitung von geometrischen Daten (Point, Polygon, Lines) Voraussetzung z.B. als Ergebnisse von Analyseanfragen.
- **Optimale Nutzerführung hinsichtlich der Darstellung raumbezogener Daten**
Je nach Fragestellung muss der Anwender entscheiden, ob eine tabellarische, chart-orientierte oder kartenbasierte Darstellung oder eine Kombination der Darstellungsformen sinnvoll ist. In Geo-Visualisierung ungeübte Nutzer benötigen hier eine adäquate Unterstützung – insbesondere bei der Wahl der geeigneten Kartendarstellung.
- **Synchronisierung aller Darstellungsarten**
Da als weitere Darstellungsform die Karte hinzukommt, müssen sich Änderungen in einer Darstellungsform synchronisiert auch in die anderen Darstellungsformen auswirken. Ein Zoomen in der Karte z. B. führt zu einem entsprechend höheren Detailgrad in der Tabelle oder im Chart bzw. viceversa.
- **Darstellung entscheidungsrelevanter Daten in der Karte**
Durch geeignete Navigationsmöglichkeiten, visuelle Hervorhebung signifikanter Details oder Identifikation statistisch signifikanter Hot Spots, Cold Spots oder Ausreißer ist die Entscheidungsunterstützung mittels Karteninformationen optimal zu gestalten.

2.2 Technologische Anforderungen

Neben den typischen technischen Anforderungen an klassische BI Systeme wie Datenintegration (ETL Prozess), OLAP Funktionalität (Cube-Generierung, Slice & Dice, Roll-up & Drill-down, Pivotierung, Sortierung ...) oder Reporting, Forecasting und Data Mining ergeben sich für ein um Geo-Funktionalität erweitertes BI System weitere technologische Anforderungen:

- Datenintegration geo-referenzierter Daten (ETL Prozess mit Geo-Funktionalität)
- Erweiterung der OLAP Funktionalität um raumbezogene Analysemöglichkeiten
- Erweiterung der Reporting Funktionalität um Kartendarstellungen

3 Architekturkonzept als Best Practice Ansatz

3.1 Auswahl der Open Source BI Lösung

Derzeit stehen unterschiedliche Open Source BI Systeme zur Verfügung, die einen Reifegrad erreicht haben, der einen Einsatz in professioneller Umgebung erlaubt.

Für dieses Forschungsprojekt wurden die drei Open Source BI Systeme Jaspersoft, Palo und Pentaho untersucht; analysiert und getestet wurde jeweils die Community Edition. Aufgrund von Performanzproblemen schied Jaspersoft frühzeitig aus; dabei ließ sich nicht abschließend klären, ob es an der im Projekt es der installierten Version lag oder gegebenenfalls ein generelles Problem besteht.

Oftmals genannte Kriterien zur Evaluierung von BI Lösungen wie z. B. Unterstützung des ETL-Prozesses, Vielzahl von Analysemöglichkeiten, Bereitstellung aller relevanten BI Tools o.ä.

(vgl. z. B. [9], [12], [13]) waren im Rahmen der Entscheidungsfindung zwischen Pentaho und Palo untergeordnet, da beide Systeme in diesem Bereich ähnliche Ergebnisse liefern bzw. vergleichbare Funktionalität anbieten:

- Die klassischen BI Funktionalitäten bzgl. Reporting und Analyse, die im Rahmen des Projektes von Relevanz sind, waren bei beiden Systemen gegeben; Möglichkeiten des Data Mining wurden vernachlässigt. Hinsichtlich Gesamtpformance –Ladezeiten, Datenintegration, Weboberfläche, Verarbeitung umfangreicher Datenmengen – waren beide Systeme gut vergleichbar.
- Über die BI Funktionalität hinaus war für die Entwicklung eigener Komponenten eine einfache Anbindung eigener Komponenten durch Standardschnittstellen (z. B. MDX via XMLA) von großer Bedeutung (vgl. z. B. [11]); auch diese Möglichkeit war bei beiden Systemen gegeben.

Von zentraler Bedeutung für die Entscheidungsfindung war jedoch die Möglichkeit, geo-referenzierte Daten beim Import in das Data Warehouse sowie bei der Analyse zu verarbeiten. Hier gibt es bereits Entwicklungen für pentaho, die eine Anbindung und Nutzung raumbezogener Daten unterstützen. [24]

Ein Lösungsansatz zur Ergänzung der klassischen BI Architektur unter Verwendung von pentaho mit Geo-Komponenten wird in [24] beschrieben. Geo-Mondrian und Geo-Kettle werden als entsprechende Erweiterungen zu Mondrian bzw. Kettle angeboten. Mit Geo-Kettle lassen sich geo-bezogene Daten beim Import verarbeiten, Geo-Mondrian wiederum erlaubt die Formulierung raumbezogener Abfragen beim Zugriff auf BI Daten (vgl. auch [5]).

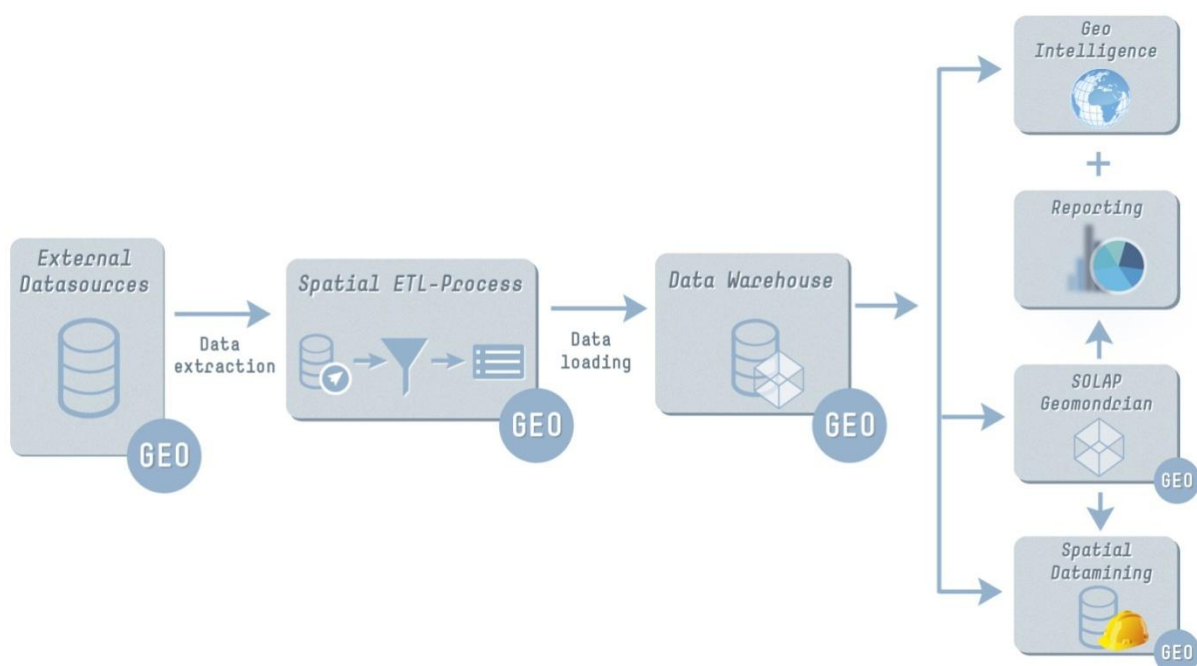


Bild 3: BI Infrastruktur auf Basis von pentaho mit Geo-Komponenten

Dieser in Bild 3 dargestellte Ansatz wurde im hier vorgestellten Vorhaben verfolgt.

3.2 Konzept einer Rich Internet Application

3.2.1 Technologische Basis der Rich Internet Application

Für exploratives Analysieren ist eine hohe Interaktivität Voraussetzung. Dies wurde in der Vergangenheit bei sogenannten RIA (Rich Internet Applications) Web-Anwendungen nahezu ausschließlich mit Adobe Flash oder MS Silverlight realisiert (vgl.z. B. [17]).

Mit der Verfügbarkeit und zunehmenden Verbreitung von HTML5 stehen neue zukunftsweisende Möglichkeiten bereit, die sich insbesondere für Karten- und Chartbasierte Visualisierungen anbieten (siehe z. B. [2]). Die technologische Grundlage dieser RIA basiert daher auf HTML 5 in Kombination mit JavaScript und der jQuery-Bibliothek.

Unter Verwendung dieser Technologien wurde eine Widget-basierte Architektur für die RIA entwickelt. Die einzelnen Komponenten und ihre Kommunikation werden nachfolgend vertiefend betrachtet.

3.2.2 Funktionalität serverseitig

Eine Postgres Datenbank mit der „Spatial“ Erweiterung Postgis dient auf unterster Ebene als Basis für Data Warehouse, für Benutzerverwaltung und insbesondere für geo-graphische Abfragen.

Ein Tomcat Server dient als Webserver; Applikationsserver binden die Anwendungsmodule wie User-Management und Geo-Mondrian ein. Geo-Mondrian stellt den Datentyp „geometry“ (Point, Polygon, Lines) zur Verfügung und erlaubt die Handhabung dieser Daten direkt in den Datencubes. [10]Die Verarbeitung der Geo-Daten erfolgt clientseitig mit direktem Zugriff auf OSM (OpenStreetMap) Daten.

Das Zusammenspiel der Komponenten wird in Bild 4 verdeutlicht.

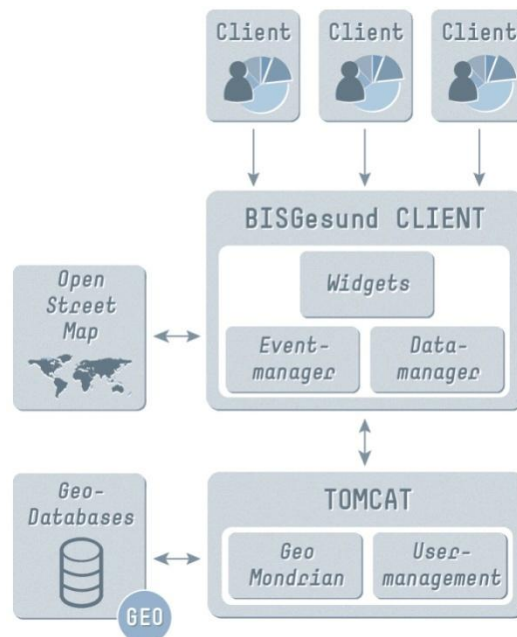


Bild 4: Client-Server Zusammenspiel

3.2.3 Funktionalität clientseitig

Clientseitig liegt der Schwerpunkt auf interaktiven OLAP-Analyse- und Auswertungsmöglichkeiten sowie der Visualisierung auf bzw. der Interaktion mit der Karte. Diese Funktionalität wird über eine Widget-Architektur realisiert. Dieses Architekturkonzept ermöglicht einen flexiblen und erweiterbaren Ansatz.

Widgets sind kleine, eigenständige wiederverwendbare Programmteile, typischerweise mit einer graphischen Repräsentation, z. B. einem eigenen Fenster und einer klar spezifizierten Aufgabe.[14],[23]Im gewählten Architekturkonzept kommunizieren Widgets über Events rein clientseitig untereinander und stellen damit eine performante sowie direktmanipulative Interaktion mit dem System sicher[3](vgl. Bild 5).

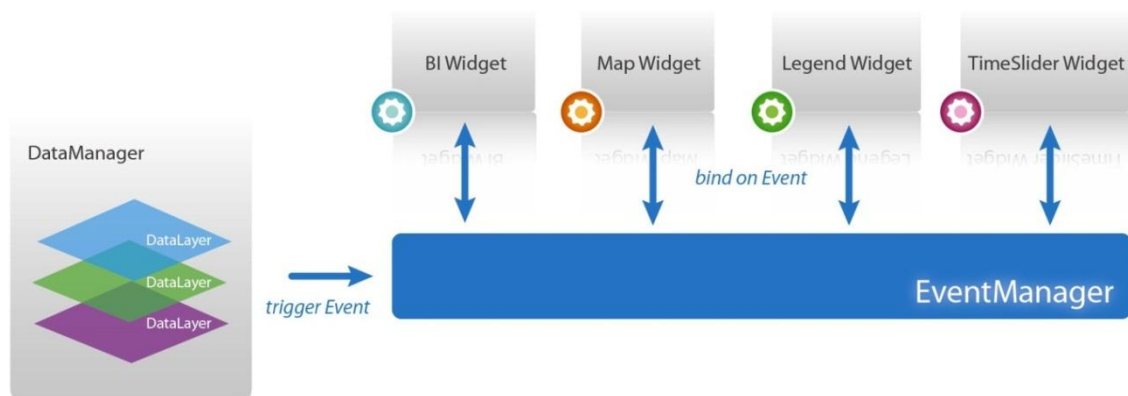


Bild 5: Widget-Architektur

Ein zentraler Client-Datencontainer optimiert die Kommunikation zwischen Client und Server und stellt zudem eine konsistente Datenhaltung in allen Widgets sicher. Widgets werden für folgende Aufgaben eingesetzt: BI OLAP, Maps, Charts, Layer, Projektmanagement sowie User-Management.

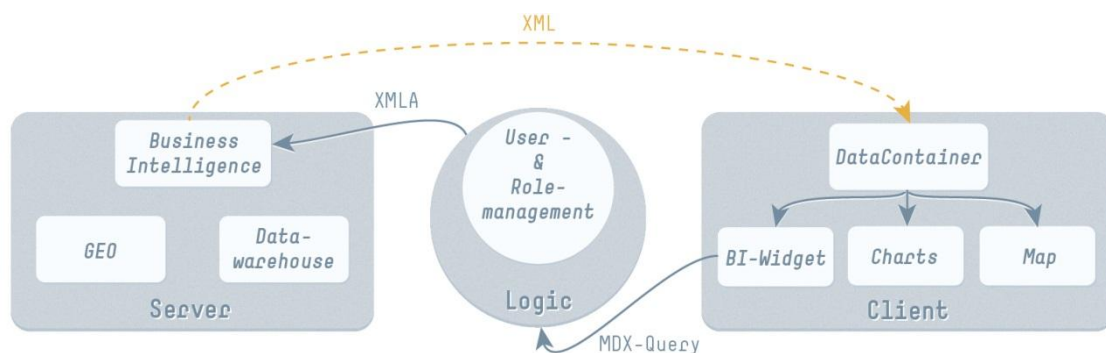


Bild 6: Client Sever Komponenten des BISGESUND Systems

3.2.4 Workflow

Die Kommunikation zwischen den Komponenten ist von zentraler Bedeutung. Bei der technologischen Realisierung wurde darauf geachtet, dass Standardabfragesprachen und allgemein übliche Kommunikationsformate eingesetzt werden.

Das BISGesund System nutzt daher Multidimensional Expressions (MDX) sowie „XML for Analysis“ (XMLA) zur Definition der Abfragen. Die Abfrageresultate werden in XML übertragen. Exemplarisch ist der Workflow zwischen den Komponenten ist durch die Pfeile in Bild 7 visualisiert.

Die Widgets formulieren – ausgelöst durch Benutzeraktionen – MDX Queries. Die Business-Logic prüft deren Zulässigkeit und stellt die Anfrage mittels XMLA an den BI Server – im Detail an die Geo-Mondrian Komponente. Das Ergebnis der Abfrage wird an den Datencontainer übermittelt (XML-Format), der die Widgets mit den aktuellen Daten versorgt. Die Anbindung an Geo-Mondrian erfolgt mit dem XMLA4JS Framework. Dieser Ablauf bildet die Basis für alle Benutzeraktionen im Web-Frontend. Klassische Datenstrukturen und geo-referenzierte Daten sowie deren Abfragen werden gleichermaßen behandelt.

Der Nutzer agiert typischerweise mit drei User Interface Komponenten: Pivot-Tabelle, Karte sowie Charts. Exploratives Analysieren bedeutet das Arbeiten mit den Daten in allen drei „Ansichten“. Manipulation auf der Karte beeinflussen daher auch direkt die Darstellungen in der Pivot-Tabelle sowie in den Charts. Der resultierende Workflow wird exemplarisch für ein Karten- sowie für eine Pivot-Tabelle-Interaktion erläutert.



Bild 7: Pan & Zoom Karteninteraktion

Durch eine Pan & Zoom Interaktion an der Karte verändert sich der Datenbereich (z. B. Europa → Deutschland) (vgl. Bild 7). Dies führt zu einer Veränderung der Bounding Boxen im Datencontainer, der – basierend auf der Veränderung – eine neue MDX Abfrage mit veränderten Bounding Boxen initiiert. Die Abfrage wird durch Geo-Mondrian verarbeitet und Daten werden – über den Datencontainer – an die anderen Widgets weitergereicht.

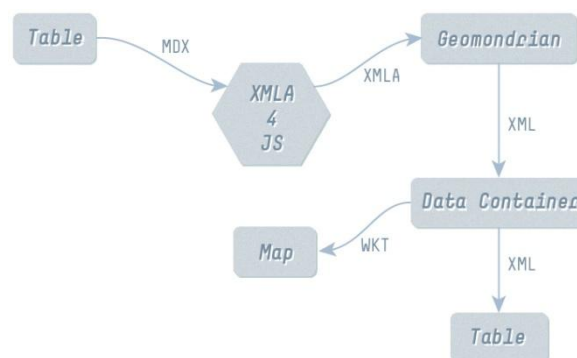


Bild 8: Ablauf einer OLAP Interaktion in einer Pivot Tabelle

Der Ablauf bei einer OLAP Interaktion in einer Pivot-Tabelle wie z. B. Drill-down oder Auswahl einer Dimension, die auf der Karte zu visualisieren ist, erfolgt analog (vgl. Bild 8). Geo-Mondrian verarbeitet die Anfrage und schickt das Ergebnis an den Datencontainer – hier explizit dargestellt. Die Kommunikation zwischen Datencontainer und Karte erfolgt für Vektordaten in dem WKT-Format.

4 Evaluation der BI Lösung

Eine erste Evaluierung des Prototyps erfolgte im Kontext „Fortbildungsverhalten bei Gesundheitsberufen“. Ziel war, einen realen Anwendungsfall zu nutzen, der möglichst viele Facetten der Anwendung abdeckt und insbesondere Daten sowie Fragestellungen mit geo-graphischem Bezug aufweist.

4.1 Anwendungsszenario Fortbildungsverhalten bei Gesundheitsfachberufen

Mit dem GKV-Modernisierungsgesetz (GMG) verpflichtet der Gesetzgeber die Vertragsärzte, sich regelmäßig fachlich fortzubilden und die Nachweise alle fünf Jahre der Kassenärztlichen Vereinigung (KV) vorzulegen. Vertragsärzten müssen regelmäßig nachweisen, dass sie sich fortgebildet haben. Neben der Fortbildungspflicht bei den Vertragsärzten existieren für zahlreiche weitere Gesundheitsfachberufe freiwillige Fortbildungsangebote.

Ziel der Analyse sind belastbare Ergebnisse, um das Fortbildungsangebot des Anwendungspartners unter Berücksichtigung der Fortbildungsqualität und der Kostenstruktur zu optimieren. Als potentielle Anwendungspartner sind in erster Linie die Veranstaltungsinits der medizinischen Fachverlage und diverse Berufsverbände zu nennen. Folgende Parameter bzw. Fragestellungen sind für die Analyse von Relevanz:

- Welche Form der Fortbildungsangebote werden genutzt? Gibt es regionale bzw. saisonale Unterschiede bei der Nutzung der Fortbildungsangebote? Gibt es beliebte bzw. unbeliebte Veranstaltungsorte? Welche Entfernungen vom Wohnort zum Veranstaltungsort werden in Kauf genommen?
- In welchen Bereichen werden welche Einnahmen erzielt?
- Welche Beziehungen zu Teilnehmer, Einreicher, Referenten, Schwerpunktthemen, einzelnen Programmpunkten lassen sich in Bezug auf geo-graphische Verteilung und zeitliche Veränderung ableiten?
- Zeigen Marketingmaßnahmen, Produktpolitik, Nachfrageaspekte Wirkungen? Falls ja, welche?

4.2 Test des Prototypen

Mit einem umfangreichen Datensatz, der von einem Fachanbieter für medizinische Kongresse zur Verfügung gestellt wurde, konnten erste Erfahrungen mit dem Prototyp gesammelt werden (vgl. Bild 9).

Der Mehrwert der Kartenvisualisierung zusätzlich zu klassischen Darstellungsformen ist offensichtlich. Bei der Darstellung in Bild 9 ging es um die Anmeldezahlen für einen Kongress und deren Verteilung auf die Bundesländer. Das Tortendiagramm als Chart zeigt zwar die relative Verteilung; die Kartendarstellung jedoch vermittelt erst einen flächenhaften Eindruck der geo-graphischen Verteilung der Anmeldezahlen.

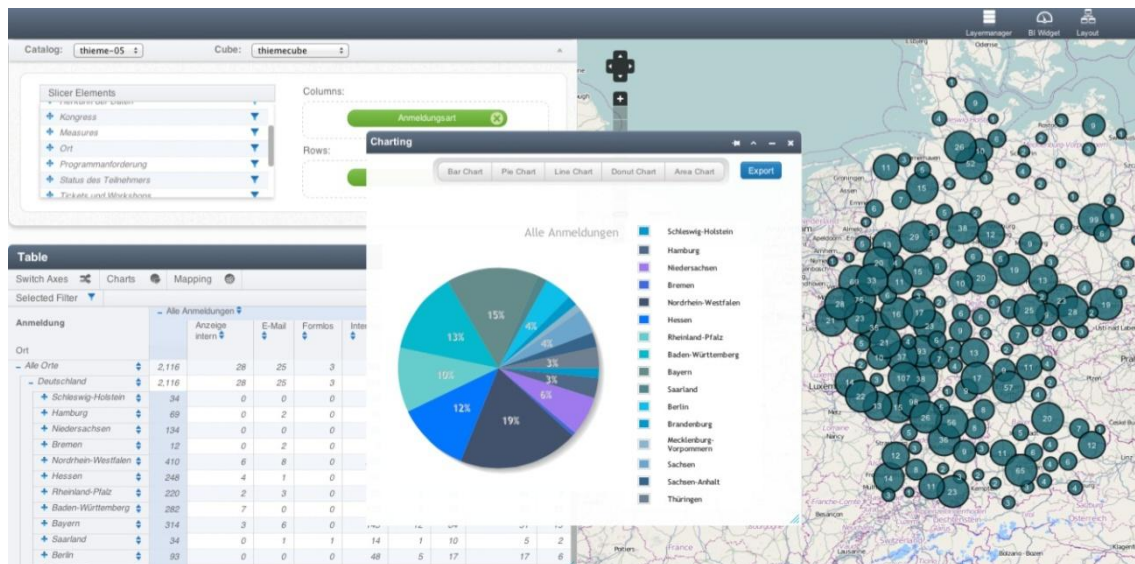


Bild 9: Screenshot des Prototyps

Der Anwender nutzt unterschiedliche Zugänge zu den Daten; es eröffnen sich neue Analyse- und Visualisierungsmöglichkeiten. Über Pivot Tabellen kann er die ihm zur Verfügung stehenden Daten insbesondere in Cubes klassisch untersuchen; Attribute ausgewählter Spalten oder Zeilen werden direkt in der Karte in gewünschter Form visualisiert. Unterschiedliche interaktive Charts lassen sich bei Bedarf hinzunehmen und Analysen vertiefen.

Bei der Realisierung des Prototyps lag ein Schwerpunkt auf der Synchronisierung der Darstellungsarten, um visuelle Vergleiche über unterschiedliche Zustände –z.B. historische/aktuelle Situationen– optimal zu gestalten. Die Kartendarstellung lässt sich quasi einfrieren und bei Bedarf mit ergänzenden Analysen vervollständigen.

4.3 Kritische Reflexion

Der vorgestellte Prototyp folgt einem Best-Practice Ansatz und soll aufzeigen, welche Möglichkeiten in der Kombination einer Open Source BI Lösung und Geovisualisierung für das Gesundheitswesen bestehen.

Dennoch sind verschiedene Fragestellungen noch offen bzw. müssen nach Abschluss des Forschungsprojekts noch geklärt werden:

- **Akzeptanz**
Erste Tests im Einsatz bei Anwendern – ein Fachanbieter für medizinische Kongresse – zeigt, dass die Anwender die neuen Visualisierungsmöglichkeiten nutzen, schätzen und als deutliche Bereicherung empfinden.
- **Datensicherheit**
Das Forschungsprojekt zielte darauf ab, dass ein entsprechendes System dem Anwender in Form eines SaaS zur Verfügung gestellt wird. Zahlreiche Diskussionen mit potenziellen Anwendern haben aber gezeigt, dass für Anwender gegebenenfalls nur eine private Cloud Lösung infrage kommt. Alles andere wird unter dem Aspekt der Datensicherheit zum jetzigen Zeitpunkt noch zu kritisch beurteilt. Aus technischer Sicht ist das Konzept für den Einsatz in einer privaten Cloud geeignet. Möchte man ein komplett abgeschottetes Intranet betreiben, sind die OSM Kartendaten noch in einem eigenen Mapsserver bereitzustellen.

- Data Mining Funktionalitäten

Der Aspekt Data Mining wurde im Rahmen des Projektes für pentaho detaillierter analysiert; die Ergebnisse der Recherche zeigen zwar entsprechende Lösungen wie z. B. Weka. In der Diskussion mit potenziellen Anwendern zeigte sich aber, dass die Anwender i.d.R. nicht über das notwendige Statistikwissen verfügen, um Data Mining Komponenten adäquat und mit genügend Fachwissen zu nutzen. Die Anwender stellten heraus, dass umfangreiche Tools zwecks Reporting und Analyse eine wesentlich höhere anwendungsorientierte Relevanz besitzen.

- Einsatzmöglichkeiten außerhalb des Gesundheitswesens

Derzeit laufen erste Gespräche mit potenziellen Anwendern im Bereich Vertrieb sowie Medieninformation bzw. -nutzung. In beiden Fällen kann erst die Berücksichtigung georeferenzierter Daten aussagekräftige Resultate liefern kann.

- Produktvermarktung

Erste Gespräche mit potenziellen Anwendern haben auch gezeigt, dass einige Anwender eher daran interessiert sind, das Tool inklusive Daten zu erwerben bzw. lizenzieren als nur das Tool selbst. Oftmals haben die Anwender Schwierigkeiten, den ETL-Prozess selbst durchzuführen und müssten hier eine entsprechende Dienstleistung einkaufen. Zum anderen sind für einige Anwender im Gesundheitswesen auch bereits eingestellte Daten, die als Daten-Stream wie z. B. Luftverschmutzung frei verfügbar sind, von großer Bedeutung, die sie mit ihren eigenen Daten verschneiden können.

5 Fazit und Ausblick

In diesem Beitrag haben wir einen technologischen Lösungsansatz vorgestellt, der eine Open Source BI Plattform mit Geovisualisierung kombiniert. Der Ansatz berücksichtigt die dargelegten Anforderungen insbesondere hinsichtlich Integration der raumbezogenen Daten in den Analyseprozess und der synchronen Visualisierung und Interaktion.

Ein erster Prototyp liegt vor, der von verschiedenen Anwendern aus dem Gesundheitsmarkt getestet wird. Die bisherigen Arbeiten mit dem Prototyp zeigen, dass das gewählte Architekturkonzept vielversprechend hinsichtlich Flexibilität und Performance ist. In diesem Prototyp kommt derzeit nur ein Anwendungsszenario zur Anwendung. In weiteren Schritten werden noch weitere Szenarien in dem Prototyp umgesetzt.

Zukünftige Arbeiten sollten die in Abschnitt 4.3 oben kritisch angesprochen Punkte weiter betrachten. Insbesondere die Nutzungsmöglichkeit von Data Mining gegebenenfalls durch vordefinierte Analyseansätze sollte untersucht werden. Weiterhin ist zu prüfen, in wie weit ETL-Prozesse mit raumbezogenen Daten für vordefinierte Anwendungsfälle vereinfacht werden können. Die Visualisierung insbesondere auf der Karte ist noch nicht erschöpfend betrachtet. Die Nutzbarkeit von weiteren Darstellungsmöglichkeiten ist zu untersuchen.

Das Projekt endete Ende 2011; zu diesem Zeitpunkt lag ein stabiler Prototyp vor, der die Basis für eine spätere Verwertung im Gesundheitsmarkt bildet.

6 Literatur

- [1] Bisgesund (2011). <http://www.bisgesund.de/> Abgerufen am 14. Dezember 2011.
- [2] Boulos, M. K., Warren, J., Gong, J., Yue, P. (2010). Web GIS in practice VIII: HTML5 and the canvas element for interactive online mapping", International Journal of Health Geographics, Vol. 9, No. 1.
- [3] Böhm, K., Mehler-Bicher, A., Fenchel, D. (2011). GeoVisualAnalytics in the Public Health Sector. In: Spatial Data Mining and Geographical Knowledge Services (ICSDM), 2011 IEEE International Conference on Issue, June 29 2011–July 1 2011 pp. 291-294.
- [4] Chamoni, P., Gluchowski, P. (2004). Integrationstrends bei Business-Intelligence-Systemen – Empirische Untersuchung auf Basis des Business Intelligence Maturity Model, Wirtschaftsinformatik 46 (2004) 2, S. 119-128.
- [5] Eichar, J. (2011) The Intersection of Geospatial and Business Intelligence, Presentation at FOSS4G2011. <http://foss4g.org/drupal/content/intersection-geospatial-and-business-intelligence>. Abgerufen am 15. Dezember 2011.
- [6] Feix, C. (2007). Bedeutung von 'Geo-Business Intelligence' und Geomarketing zur Entscheidungsunterstützung unternehmerischer Planungsprozesse im Kontext wirtschaftlicher Liberalisierung, FU-Berlin, http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000003334. Abgerufen am 12. September 2011.
- [7] Fisher, D. (2007). Hotmap. Looking at Geographic Attention, http://research.microsoft.com/pubs/69446/fisher_infovis_hotmap.pdf. Abgerufen am 12. September 2011.
- [8] Fraunhofer (2007). http://www.iuk.fraunhofer.de/index2.html?Dok_ID=74&Sp=1&MID=1281&PHPSESSID=10f90dbc635eafd39a868b565e4067ed. Abgerufen am 16. August 2011.
- [9] Fritsch, W., Wilms, J.: Business Intelligence: Wer richtig auswählt, spart Zeit und Geld, <http://www.crn.de/software/artikel-84671-3.html>, Abgerufen am 14. Dezember 2011.
- [10] Geosoa (2011). http://geo-soa.scg.ulaval.ca/en/index.php?module=pagemaster&PAGE_user_op=view_page&PAGE_id=19. Abgerufen am 6. September 2011.
- [11] Gluchowski, P., Gabriel, R., Dittmar, C. (2008) Management Support Systeme und Business Intelligence. Computergestützte Informationssysteme für Fach- und Führungskräfte, 2. Auflage, Springer, Heidelberg.
- [12] Gluchowski, P., Schieder, C. (2009) Open Source Business Intelligence: Acht quelloffene Werkzeuge für Reporting, OLAP und Data Mining im Vergleich, BARC, ISBN 978-3-937818-41-2.
- [13] IT Novum (2009). Open Source Business Intelligence: Ein Vergleich der Open Source BI Lösungen JasperSoft, Pala und Pentaho, <http://www.it-novum.com/download/downloads/whitepaper-open-source-business-intelligence.html> Abgerufen am 15. Dezember 2011.
- [14] IT Wissen (2011): Widget Definition, <http://www.itwissen.info/definition/lexikon/Widget-widget.html> Abgerufen am 16. Dezember 2011.

- [15] Jaspersoft. (2009). RightScale Joins Forces with Jaspersoft, Talend and Vertica to Deliver Complete Business Intelligence on the Cloud. <http://www.jaspersoft.com/press/rightscale-joins-forces-with-jaspersoft-talend-and-vertica-deliver-complete-business-intelligence>. Abgerufen am 12. September 2011.
- [16] Jaspersoft. (2011). <http://www.jaspersoft.com>. Abgerufen am 12. Spetember 2011.
- [17] Jern, M., Thygesen, L., Brezzi, M. (2009). A Web-enabled Geovisual Analytics Tool applied to OECD Regional Data. In: Reviewed Proceedings in Eurographics 2009, München.
- [18] Kemper, H.-G., Mehanna, W., Unger, C. (2006). Business Intelligence: Grundlagen und praktische Anwendungen, 2. Auflage, Vieweg, Wiesbaden.
- [19] Kohlstock, P. (2010). Kartographie– eine Einführung, UTB, Stuttgart. 2. Aufl., März 2010.
- [20] Kraak, M.-J. (2007). <http://www.itc.nl/personal/kraak/publications/2007/2007-xian-Geo-visual%20Analytics-autumn.pdf> Abgerufen am 12. September 2011.
- [21] Palo. (2011). <http://www.palo.net/en>. Abgerufen am 12. September.
- [22] Pentaho. (2011). <http://www.pentaho.com>. Abgerufen am 12. September 2011.
- [23] Supercharged JavaScript Graphics: with HTML5 Canvas, SVG, JQuery, and More. (2011) O'Reilly Media.
- [24] Tomaszewski, B. (2009). Emerging Applications and Challenges for Geovisual Analytics Research, <http://rc.rit.edu/docs/rcss-tomas-presentation.pdf> Abgerufen am 6. September 2011.
- [25] Weckbecker, A. (2010) Geovisual Analytics. Visualisierungstechniken, Einsatzgebiete und Beispiele in Online-Anwendungen, VDM Verlag Dr. Müller.